# The 3-D structure of a folate-dependent dehydrogenase/cyclohydrolase bifunctional enzyme at 1.5 Å resolution

Marc Allaire[1,2], Yunge Li[1,2], Robert E MacKenzie[2,3] and Miroslaw Cygler[1,2]*

**Background:** The interconversion of two major folate one-carbon donors occurs through the sequential activities of NAD(P)-dependent methylene[$H_4$]folate dehydrogenase (D) and methenyl[$H_4$]folate cyclohydrolase (C). These activities often coexist as part of a multifunctional enzyme and there are several lines of evidence suggesting that their substrates bind at overlapping sites. Little is known, however, about the nature of this site or the identity of the active-site residues for this enzyme family.

**Results:** We have determined, to 1.5 Å resolution, the structure of a dimer of the D/C domain of the human trifunctional cytosolic enzyme with bound NADP cofactor, using the MAD technique. The D/C subunit is composed of two α/β domains that assemble to form a wide cleft. The cleft walls are lined with highly conserved residues and NADP is bound along one wall. The NADP-binding domain has a Rossmann fold, characterized by a modified diphosphate-binding loop fingerprint – GXSXXXG. Dimerization occurs by antiparallel interaction of two NADP-binding domains. Superposition of the two subunits indicates domain motion occurs about a well-defined hinge region.

**Conclusions:** Analysis of the structure suggests strongly that folate-binding sites for both activities are within the cleft, providing direct support for the proposed overlapping site model. The orientation of the nicotinamide ring suggests that in the dehydrogenase-catalyzed reaction hydride transfer occurs to the pro-*R* side of the ring. The identity of the cyclohydrolase active site is not obvious. We propose that a conserved motif – Tyr52-X-X-X-Lys56 – and/or a Ser49-Gln100-Pro102 triplet have a role in this activity.

Addresses: [1]Biotechnology Research Institute, NRC, 6100 Royalmount Avenue, Montréal, Québec, Canada H4P 2R2, [2]Montréal Joint Centre for Structural Biology and [3]Department of Biochemistry, McGill University, 3655 Drummond St, Montréal, Québec, Canada HG 1Y6.

*Corresponding author.
E-mail: mirek.cygler@bri.nrc.ca

**Key words:** bifunctional, channeling, cyclohydrolase, dehydrogenase, folate

Received: **15 September 1997**
Revisions requested: **3 November 1997**
Revisions received: **1 December 1997**
Accepted: **5 January 1998**

**Structure** 15 February 1998, **6**:173–182
http://biomednet.com/elecref/0969212600600173



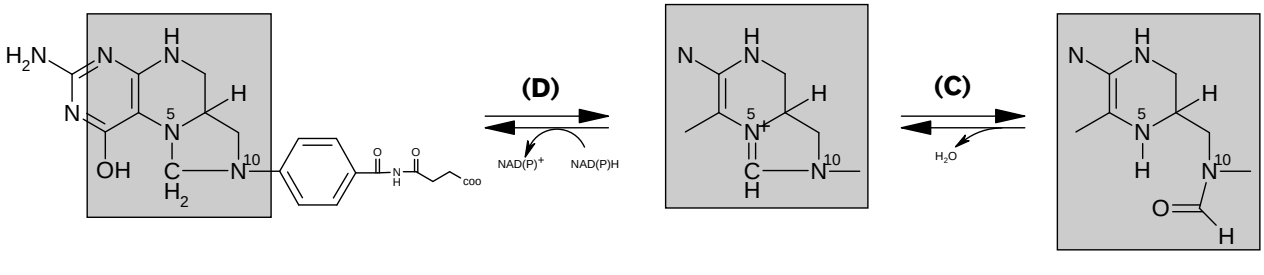
## Introduction

The recognition that some aspects of vascular disease and certain birth defects are affected by the folate nutrition status of individuals [1,2] has renewed interest in the basic metabolic pathways that support one-carbon transfers, particularly with respect to methylation reactions. The major one-carbon folate donors are $N^5$-methyl[$H_4$]folate, a methyl donor in methionine biosynthesis, $N^5$,$N^{10}$-methylene[$H_4$]folate, which provides the methyl group for thymidylate synthesis, and $N^{10}$-formyl[$H_4$]folate, a formyl group donor in purine and formylmethionine-tRNA biosyntheses. The oxidation of the metabolic intermediate $N^5$,$N^{10}$-methylene[$H_4$]folate to $N^5$,$N^{10}$-methenyl[$H_4$]-folate requires the enzyme methylene[$H_4$]folate dehydrogenase (D). Although the methenyl-product is unstable at neutral pH and hydrolyzes quite readily, most dehydrogenases have an associated methenyl[$H_4$]folate cyclohydrolase (C) activity which catalyzes its hydrolysis to $N^{10}$-formyl[$H_4$]folate (Figure 1).

The D/C enzymes are dimers and occur as bifunctional enzymes or as bifunctional domains of a trifunctional D/C/S enzyme that has a 67 kDa C-terminal domain with formyl[$H_4$]folate synthetase (S) activity [3]. The known monofunctional dehydrogenases are also dimeric and generally use NAD as cofactor, whereas the bifunctional domains are predominantly NADP dependent. A notable exception to this is the mammalian mitochondrial D/C enzyme that uses NAD and requires phosphate and $Mg^{2+}$ for activity.

Intriguingly, the bifunctional D/C enzymes have molecular masses similar to those of the monofunctional D enzymes (~34 kDa). Together this observation and the relatively small size of the bifunctional enzymes suggest that there is a common binding site for the dehydrogenase and cyclohydrolase substrates. Chemical modification studies of the D/C enzymes also indicated that the dehydrogenase and cyclohydrolase substrate-binding sites are at least in close proximity and might actually overlap [4–7]. Kinetic properties of the enzymes, such as partial (50%) channeling of the methenyl[$H_4$]folate product of the dehydrogenase to the cyclohydrolase [8–10] and partial inhibition of the hydrolysis of added methenyl[$H_4$]folate by the nicotinamide moiety of NADP [11], also support such an interpretation. Binding studies

**Figure 1**




Reactions catalyzed by the dehydrogenase (D) and cyclohydrolase (C) activities of DC301.

indicate the presence of only one folate site per polypeptide and, in conjunction with other data, recently led us to propose that a single folate-binding site provides particular advantage in transforming formyl- to methylene[$H_4$]-folate [11,12].

About 20 enzymes are known to belong to the D and D/C family, however, little is known about the details of their catalytic mechanism, the location of the active site and the identity of the catalytic residues. Even the type and location, within the primary sequences, of the NADP-binding domain was uncertain. The characteristic GXGXXG fingerprint sequence of the expected Rossmann fold [13,14] is not present among the conserved residues in the aligned sequences. In order to address these questions and, in particular, to understand the details of the catalytic mechanism, to unravel the molecular basis for the channeling between the two activities and identify the determinants that differentiate the mono- from the bifunctional enzymes, we have undertaken structural investigations of an enzyme belonging to this family. Here, we present the first three-dimensional structure of such an enzyme, the dehydrogenase/cyclohydrolase domain of the human cytosolic trifunctional enzyme [10] with a bound NADP cofactor. This structure has allowed us to identify the NADP-binding domain, propose the location of substrate-binding sites and suggest the residues that are probably involved in the enzymatic activity.

## Results

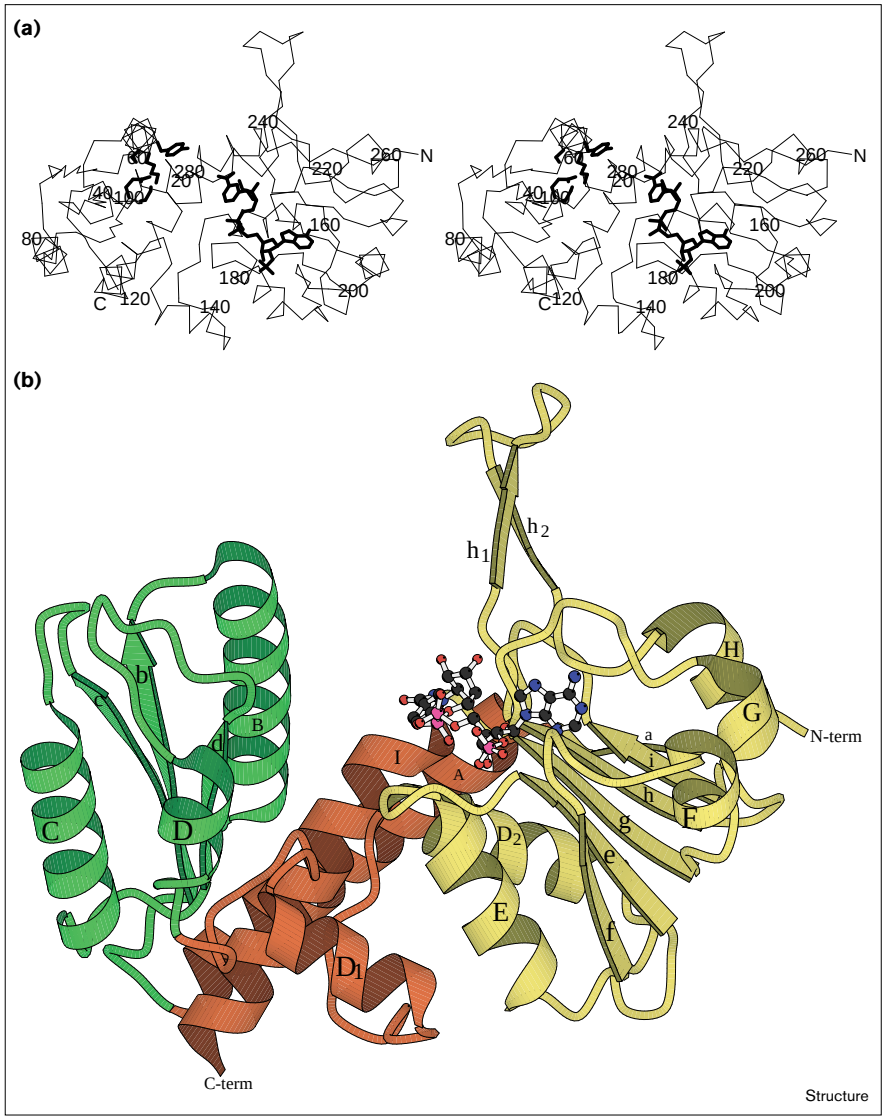
### Overall structure of the monomer

The D/C fragment of the human trifunctional D/C/S enzyme, containing residues 2–301 and referred to as DC301, is globular and has overall dimensions 35 Å × 40 Å × 55 Å. The three-dimensional structure of DC301 is shown in Figure 2. All residues are well ordered, except for the five C-terminal residues in both the crystallographically independent molecules (A and B), and the 241–250 loop, which is disordered in molecule A and not visible in the electron-density map, even at 1.5 Å resolution. DC301 is made up of two domains that are connected by two long α helices, A and I. The arrangement of these domains onto the helices creates a large cleft between them (Figure 3). The N-terminal domain, which includes residues 31–119, has an α/β fold and is composed of a three-stranded parallel β sheet (strands b, c and d) and three α helices (B, C and D). The helices cover both sides of the sheet and are nearly parallel to the strands (Figure 2). The C-terminal domain includes residues 146–277 and, in addition, several residues from the N terminus of the protein. This domain contains a six-stranded parallel β sheet (strands a, e, f, g, h and i), four α helices ($D_2$, E, F and H) and one $3_{10}$ helix (G), with the helices packing against the sides of the sheet (Figure 2). The N-terminal residues 2–8 form the first strand (a) of the β sheet. In molecule B, residues 241–250 are well defined and form a β hairpin (strands $h_1$–$h_2$), which extends away from the rest of the protein (Figure 2) and makes relatively few contacts with the rest of DC301 (see below).

### Dimer formation

The D/C/S trifunctional enzyme exists in solution as a dimer [10]. The DC301 fragment maintains the ability to form dimers and is observed in solution only in a dimeric form, as indicated by the retention time on a size-exclusion column [10] and confirmed by dynamic light scattering experiments (MA, YI and MC, unpublished data). The DC301 molecules also form tightly bound dimers within the crystal. The two crystallographically independent monomers are related by an approximate non-crystallographic twofold symmetry. The association of the two molecules into a dimer occurs through their C-terminal NADP-binding domains, mainly through the interactions of their six-stranded parallel β sheets. This involves an antiparallel interaction of two β strands (strands f) and leads to the formation of an extended, 12-stranded,

**Figure 2**



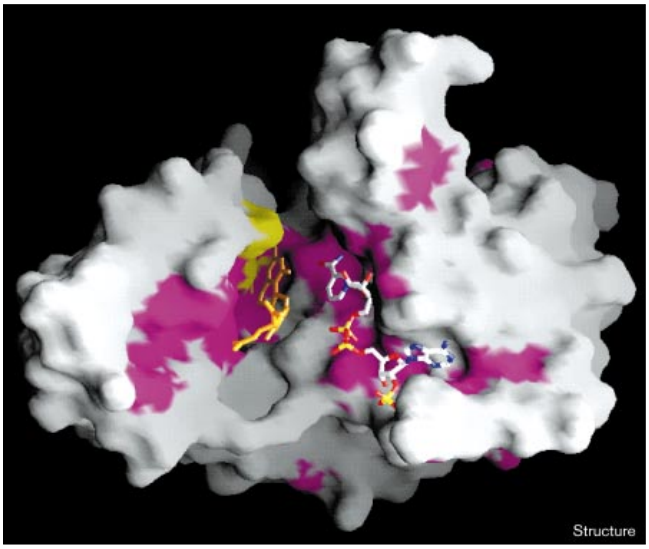

The overall structure of DC301 monomer. **(a)** Stereoview of the Cα trace with several conserved residues, discussed in the text, and NADP shown in bold stick representation. **(b)** Schematic representation of the secondary structure of the DC301 monomer. The N-terminal domain is colored green, the C-terminal NADP-binding domain is colored yellow and the connecting helices are shown in orange; the NADP molecule is shown in ball-and-stick representation with atoms colored according to type – carbon, black; oxygen, red; nitrogen, blue; and phosphorus, purple. (The Figure was prepared with MOLSCRIPT [47] and Raster3D [48].)

twisted β sheet (Figure 4). The surface area buried by this association on each molecule is 910 Å$^2$, which corresponds to ~8% of the total surface of DC301. Several hydrogen bonds are formed between the monomers during dimer formation. Four hydrogen bonds occur between the NH and CO backbone groups of residues 192 and 194 from strand f and counterpart residues 194 and 192 from the second molecule. There are also hydrogen bonds between the hydroxyl groups of Thr193 and its counterpart, as well as between Thr191 from one molecule and Thr199 from the other. Other segments of both molecules also participate in the dimer formation through van der Waals contacts; in particular, α helices $D_1$ from both molecules are brought into a close contact through their Gly133-containing sides (the distance between their Cα atoms is 3.4 Å). Two salt bridges are formed, between Arg137 of one molecule and Asp183 of the other. Several water molecules are buried in the interface and provide bridging hydrogen bonds.

### NADP-binding domain and the cofactor-binding mode

The NADP cofactor lies along one side of the cleft formed between the two domains of each DC301 monomer and makes contacts with only the C-terminal domain (Figure 3). An automated search with the DALI procedure [15] showed that this domain matches reasonably well a large number of dinucleotide-binding domains with a Rossmann fold. The best fit was obtained with D-lactate dehydrogenase (pdb code 2DLD); there are 116 structurally equivalent Cα atoms between the two molecules, encompassing five strands (b–c–d–e–f) and five helices ($D_2$–E–F–G–H) of DC301. Their superposition gave a

**Figure 3**




Molecular surface of DC301 monomer (calculated with GRASP [49]), colored white, except for the surface area associated with residues conserved in > 90% of the 15 related sequences (leucine, isoleucine and valine were grouped for this calculation) which is colored magenta. The surface area corresponding to Tyr52 and Lys56 is colored yellow. The molecule of NADP is shown as observed in the crystal structure (on the right) and a molecule of $N^5$,$N^{10}$-methylene[$H_4$]folate (orange) was modeled on the left side of the cleft.

**Figure 4**


Association of two DC301 monomers into a dimer. The two monomers are shown in different shades of gray.

root mean square (rms) deviation of 2.1 Å. The β strand (a) of DC301 at the N terminus of the protein seems to be an extension of the regular Rossmann fold. As in other NAD(P)-binding domains, the NADP molecule in DC301 is positioned on the top of the domain, over the C-terminal ends of the β strands, and has an extended conformation. The ribose rings of the nicotinamide and adenine moieties of NADP display the C2′-endo puckering in both crystallographically independent molecules.

The residues $G^{172}$RSKIVG form a loop connecting β strand e and α helix E, and they constitute the dinucleotide-binding loop in DC301. The amino acid sequence of this loop does not conform to the classical GXGXXG fingerprint pattern [13,14]. Superposition of the dinucleotide-binding domain of DC301 onto the dinucleotide-binding domain of D-lactate dehydrogenase, however, shows that the GXGXXG motif of the latter aligns structurally with residues $G^{172}RS^{174}KIVG^{178}$ of DC301. Indeed, other fingerprint patterns have been noted in recent years (GXXGXXG, malate dehydrogenases [16]; GXXXGXG, steroid dehydrogenases [17]; GXXXXXSXA, enoyl-acyl carrier protein reductase [18]). In DC301, there is a hydrogen bond between the hydroxyl group of Ser174 and one of the 5′-phosphate oxygen atoms of the nicotinamide monophosphate moiety of NADP (Figure 4). The minimal sequence requirement of the βα loop that would satisfy the observed network of hydrogen bonds with a dinucleotide is GXSXXXG, characteristic for this enzyme family, thus adding to the known fingerprint patterns for dinucleotide-binding domains. Interestingly, although most D/C sequences contain this pattern, the *folD* gene product of *Mycoplasma genitalium* contains the classic GXGXXG sequence [19].

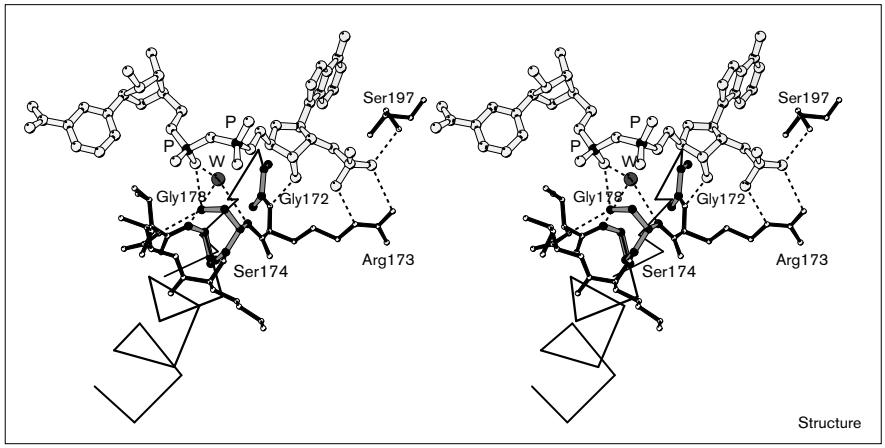
The DC301 structure clearly shows why there is a requirement for the 2′-phosphate group in the cofactor. This phosphate group is tightly bound by the enzyme. It makes multiple hydrogen bonds (Figure 5) and, in particular, one hydrogen bond to Ser197 and two to Arg173. The latter residue also provides charge stabilization.

### Comparison of the two independent molecules

The individual domains in the two molecules are nearly identical, with the exception of loop 241–250; superposition of their Cα atoms gives a rms deviation of 0.33 Å for both the N- and C-terminal domain. The relative disposition of the domains in the two independent molecules of DC301 is, however, somewhat different (Figure 6) as

**Figure 5**

Stereoview of the diphosphate-binding loop, showing binding of NADP to DC301. NADP is shown in ball-and-stick representation (light grey); the trace of residues 168–188 is shown as a continuous thick line, residues 172–178 are shown in ball-and-stick form, with Gly172, Ser174 and Gly178 indicated by thicker gray lines. Hydrogen bonds between phosphate groups of NADP and Ser174 and a water molecule (W) are shown in dashed lines.

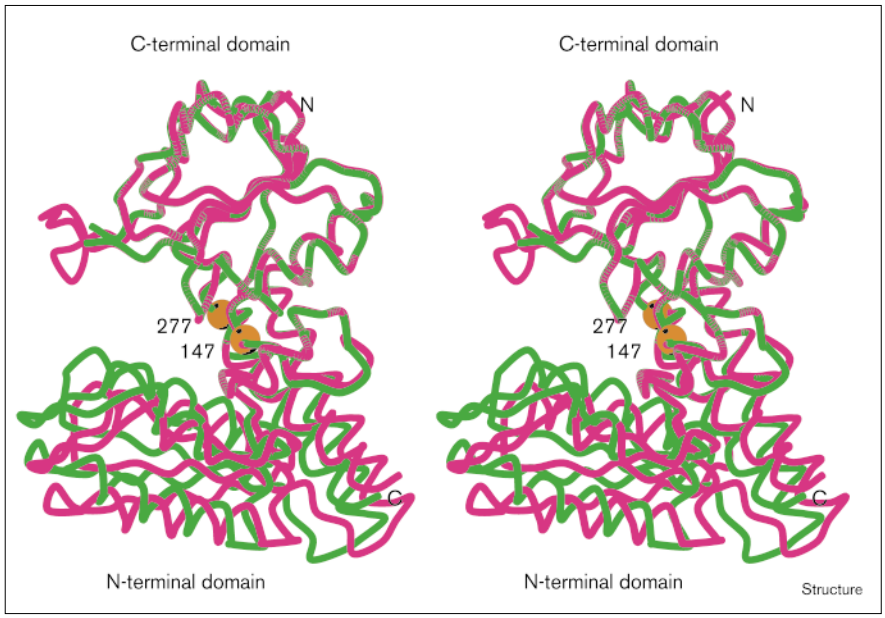



exemplified by a significantly larger rms deviation for the global superposition (1.32 Å). This difference in the relative orientation of the two domains can be described as a rigid-body motion of the N-terminal domain relative to the rest of the dimer around two hinge regions — one located within the loop connecting helices $D_1$ and $D_2$ and the other at the N-terminal end of helix I. The latter region corresponds to one of the two highly conserved blocks of sequences. The magnitude of the rigid-body movement of the N-terminal domain reaches 5.7 Å at residue 106 (Cα–Cα distance), the farthest from the hinge region. This difference in relative position and orientation of the two domains indicates the flexibility of the molecule and suggests a potential for adaptation of the size of the binding cleft to the shape of substrates.

There is also a local difference between the molecules in the dimer, namely residues 239–252 are well ordered only in molecule B, whereas there is no electron density for the equivalent sequence in molecule A. This observed variation in the mobility of the segment is most probably a reflection of different local environments of the two molecules in the crystal. The well-ordered hairpin extends away from the parent molecule and is stabilized through favorable contacts with a symmetry-related molecule; in particular, residues Asp244–Pro247, at the tip of the hairpin, form hydrogen bonds between their carbonyl groups and polar sidechains of the symmetry-related molecule. Additionally, several sidechain groups of hairpin residues interact with the other molecule through bridging waters molecules. Arrangement of this loop in the second

**Figure 6**

Stereoview of the superposition of the two crystallographically independent molecules forming the DC301 dimer. The superposition is based on the Cα atoms of the NADP-binding domain. The two molecules are shown in different colors in a ribbon representation. The orange spheres indicate the centers of the hinge regions.



molecule into the same conformation is not possible in the crystal as it would clash with the surrounding molecules.

## Discussion

Both domains of DC301 belong to known folds. The 90-residue long N-terminal domain has a typical α/β fold with a parallel β sheet and helices on both sides. A search with DALI [15] showed that it is similar to the 85 N-terminal residues of the first domain (half of the domain) of the D-ribose-binding protein (pdb code 2DRI), a representative of a large family of sugar-binding proteins. The two structures can be superimposed with a rms deviation of 2.5 Å for the Cα atoms. The C-terminal NADP-binding domain has a fold typical for dinucleotide-binding domains. Yet this fold was not recognized on the basis of its sequence by threading algorithms [20,21].

Although the individual domains of DC301 have previously observed folds, the relative disposition of these two domains, creating a deep and wide cleft, appears unique. The overall architecture of DC301 bears a certain resemblance to the family of medium-chain alcohol dehydrogenases and reductases that are also composed of a N-terminal catalytic domain and a C-terminal dinucleotide-binding domain [22].

### Overlapping substrate-binding sites

The hypothesis of the overlapping dehydrogenase and cyclohydrolase substrate-binding sites was originally suggested on the basis of the results obtained by chemical modification studies. It was shown that the modified enzyme had a coincident loss of dehydrogenase and cyclohydrolase activities, whereas the addition of folate and NADP protects the enzyme against inactivation [4–7]. The similar size of the monofunctional dehydrogenase and the bifunctional dehydrogenase/cyclohydrolase further supported this hypothesis. The kinetic observations of the inhibition of the cyclohydrolase activity by the nicotinamide moiety [11], the single folate-binding site per polypeptide [12] and the partial channeling of the $N^5$,$N^{10}$-methenyl[$H_4$]folate intermediate [8–10] all argue in favor of an overlapping site.

Analysis of the crystal structure of DC301 provides further, and strong, support for the overlapping substrate-binding site hypothesis. An important clue comes from the analysis of sequences from the D/C family. Previous alignments and analyses revealed a pattern of conserved residues, concentrated in two distinct regions: 78–102 and 272–280 (Figure 2a) [23–25]. Following the latest publication of a sequence alignment for D/C enzymes [26], we have mapped to the protein surface the residues that are conserved in 90% of the sequences from enzymes of this family (Figure 3). The area associated with the highly conserved residues maps predominantly to the large cleft between the two domains.

The binding of NADP within this cleft, with the C4 carbon of the nicotinamide ring near the bottom of the cleft, indicates that the binding of the substrate of the dehydrogenase reaction is also located within the cleft. Docking experiments indicate that the size of the cavity can accommodate only one [$H_4$]folate derivative. Although other depressions can be seen on the molecular surface, they are not large enough to bind an entire molecule of the substrate $N^5$,$N^{10}$-methenyl[$H_4$]folate and do not contain clusters of conserved residues. All these observations — the clustering of conserved residues in the cleft area, no obvious folate-binding site outside of the cleft and the size of the cleft being able to accommodate, in addition to the NADP, only one folate molecule — suggest that the folate-binding sites for the two activities either fully or partially overlap within the cleft.

### Enzymatic activities

#### *Dehydrogenase*

The crystal structure of DC301 provides insight into the catalytic mechanism of this enzyme family. The dehydrogenase-catalysed reaction occurs via a hydride transfer between $N^5$,$N^{10}$-methylene[$H_4$]folate and NADP. In general, the hydride can be transferred from the substrate either to the pro-*R* or the pro-*S* side of the nicotinamide ring. In the crystal, the nicotinamide ring is turned with its pro-*S* side toward the protein and makes contacts with Thr148, Val177, C236, Ile238, Gly276 and Thr279. The pro-*R* side is exposed to the solvent within the cleft and it is this side that would accept the transferred hydride, consistent with the observed stereochemistry [27].

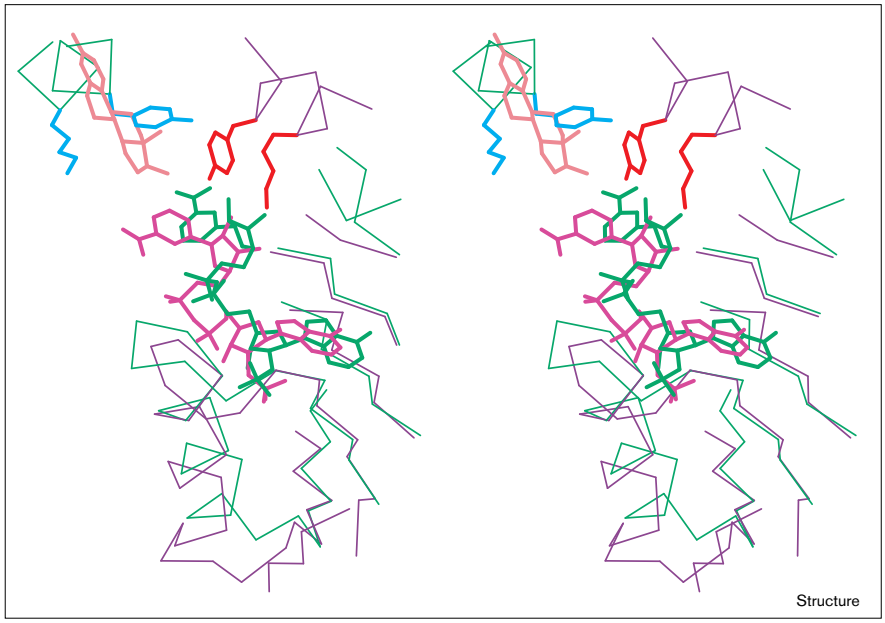
#### *Cyclohydrolase*

The predicted location of the cyclohydrolase substrate, on the basis of the overlapping substrate-binding site hypothesis, helped us to identify potential candidates for the active site of the cyclohydrolase reaction. In the vicinity of the predicted position of the substrate, we have found a $Y^{52}XXXK^{56}$ motif located on the α-helix B (Figure 3). A YXXXK motif has been postulated to form part of the catalytic site in another protein family, the short-chain alcohol dehydrogenases [28,29]. Their role in the catalysis is now better understood from the three-dimensional structure of 17β-hydroxysteroid dehydrogenase (17βHSD) in complex with estradiol and NADP [30]. In the proposed catalytic mechanism, the sidechain of this tyrosine residue is involved in the proton transfer with the substrate ketone oxygen, concomitant with the hydride transfer between NADP and the ketone carbon atom.

There are, however, specific differences between alcohol dehydrogenases and [$H_4$]folate dehydrogenases, evident in a comparison of 17βHSD and DC301 — different location of the catalytic domain relative to the NAD(P)-binding domain and different disposition of the nicotinamide ring and the YXXXK motif (Figure 7). 17βHSD catalyzes a hydride and a proton transfer in a concerted reaction. For

**Figure 7**

Stereoview of a close-up of the NADP-binding site and YXXXK motif in DC301 and 17β-hydroxysteroid dehydrogenase (17βHSD). The superposition is based on the common β strands of the NADP-binding domains. Only parts of the NADP-binding domain are shown (as a Cα trace). YXXXK motifs, NADP and estradiol, which is bound to 17βHSD, are shown in stick representation. The DC301 trace and its bound NADP moiety are shown in green, Tyr52 and Lys56 are in cyan; the trace of 17βHSD and its bound NADP moiety are shown in magenta, its Tyr155 and Lys159 in red and estradiol in pale pink. The NADP moieties are bound by the respective proteins in such a way that their nicotinamide rings are positioned to accept hydrogen atoms on opposite faces. The locations of the N-terminal domains in these two proteins relative to their NADP-binding domains are different, placing their YXXXK motifs differently with regard to the NADP.

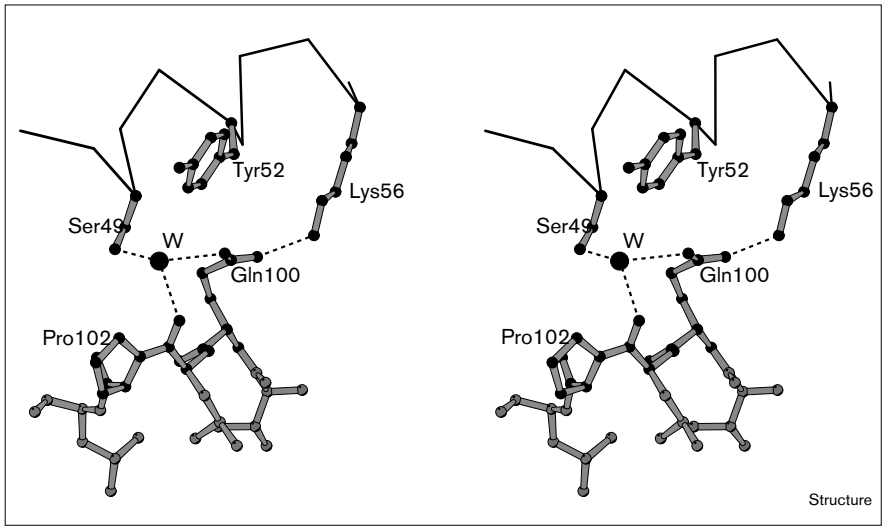



the reaction catalyzed by [$H_4$]folate dehydrogenase, in which the methylene group of the folate substrate bridges two nitrogen atoms, only the hydride transfer with the nicotinamide ring is necessary. A proton transfer is, however, required in the subsequent step of the reversible hydrolysis of the methenyl[$H_4$]folate to formyl[$H_4$]folate.

Although the evidence at this point is rather circumstantial, we postulate that the YXXXK motif plays a role in the cyclohydrolase step. The YXXXK residues are fully conserved throughout all of the D/C sequences. In contrast, the yeast monofunctional dehydrogenase [31,32], which lacks the cyclohydrolase activity but clearly belongs to the same protein family, has a threonine residue instead of a lysine. In addition to the YxxxK motif, three nearby residues, Ser49, Gln100 and Pro102, may also have a role in catalysis. They form a network of hydrogen bonds which also includes a water molecule. This water molecule is hydrogen bonded to the hydroxyl group of Ser49, the sidechain of Gln100 and the backbone carbonyl group of Leu101. The sidechain of Gln100 is also hydrogen bonded to the ε-amino group of Lys56 (Figure 8). Leu101

**Figure 8**

Hydrogen-bonding network near the YXXXK motif. The trace of residues 46–56 is shown as thick lines. Residues 99–103, as well as the sidechain of residues 49, 52 and 56 are shown in ball-and-stick representation. The water molecule is labeled with the letter W. Hydrogen bonds are shown in dashed lines.



is followed by Pro102, one of the two proline residues in *cis* conformation within the protein. The constellation of Ser/Gln/Pro residues are fully conserved throughout all of the D/C sequences, but not in the yeast monofunctional dehydrogenase.

#### Channeling of the substrate

The crystal structure of DC301 provides insight into the kinetically observed partial channeling of the methenyl intermediate from the dehydrogenase to the cyclohydrolase reaction. Our substrate-docking experiments with the DC301 structure indicate that the active site of the dehydrogenase, located at the C4 carbon of the nicotinamide ring, and the putative active site of the cyclohydrolase, located on the N-terminal domain, are far enough apart that some movement would be required to bring the $N^5$,$N^{10}$-methenyl[$H_4$]folate intermediate from one active site to the other. This could occur by the movement of folate intermediate. Alternatively, movement of the N-terminal domain could bring the cyclohydrolase catalytic residues closer to the methenyl group. Indeed, the crystal structure of DC301 indicates some degree of inter-domain flexibility. The fact that kinetic channeling from the dehydrogenase to the cyclohydrolase is only partial could be explained by a potential for the methenyl intermediate to dissociate during the movement to contact cyclohydrolase catalytic residues within the cleft.

*In vivo*, DC301 is part of a trifunctional D/C/S enzyme, in which the D/C domain has a large synthetase domain at its C-terminus. Demonstration of direct channeling of intermediate between D/C and S domains has not been shown, although the overall reverse reaction has been demonstrated in coupled kinetic experiments [33] as well as *in vivo* in yeast [34,35]. By extrapolation from the location of the C termini in the DC301 dimer, we predict that any functional interactions between active sites of the D/C and S domains would most likely involve domains from different chains, rather than from the same chain.

## Biological implications

**Folates are utilized in cells as cofactors and as carriers of one-carbon units in various metabolic reactions. $N^5$,$N^{10}$-methylene[$H_4$]folate (reduced form of folate) is either used directly in the synthesis of thymidylate or oxidized for use in other reactions. Many eukaryotic folate-dependent enzymes are multifunctional. The reversible conversion of methylene[$H_4$]folate to $N^{10}$-formate[$H_4$]folate is catalyzed by two enzymatic activities: methylene[$H_4$]folate dehydrogenase (D) converts $N^5$,$N^{10}$-methylene[$H_4$]folate to $N^5$,$N^{10}$-methenyl[$H_4$]-folate and methenyl[$H_4$]folate cyclohydrolase (C) converts $N^5$,$N^{10}$-methenyl[$H_4$]folate to $N^{10}$-formyl[$H_4$]-folate. In higher organisms, these two activities most often coexist as part of a multifunctional enzyme. In the cytosol, they are covalently linked to a third activity, formyl[$H_4$]folate synthetase (S, converts the $N^{10}$-formyl-[$H_4$]folate to [$H_4$]folate and formate). The dehydrogenase activity of the cytosolic trifunctional (D/C/S) enzyme requires NADP as a cofactor, whereas the mitochondrial enzyme (D/C) depends on NAD.**

**It has been postulated that both the dehydrogenase and cyclohydrolase substrates bind at a single site; however, little was known about the nature of this site. Furthermore, neither the residues involved in catalysis nor the NAD(P) cofactor binding site have been identified. The crystal structure of DC301, the D/C domain (residues 2–301) of the human cytosolic trifunctional enzyme, is a prototype for other dehydrogenases/cyclohydrolases and provides answers to some of these questions. The overall structure consists of two domains that assemble to form a wide cleft. The NADP cofactor binds to the C-terminal domain and lines one wall of this cleft. The NADP-binding domain has a Rossmann fold, which was not previously recognized by sequence analysis alone. The 2′-phosphate of NADP forms several hydrogen bonds and charge interactions to the enzyme, suggesting a reason for the lack of activity in the presence of NAD. Residues that are highly conserved within this protein family map predominantly to this cleft that is large enough to accommodate, in addition to NADP, only one folate molecule. The shape of the cleft and the location of conserved residues within it, provide direct support for the proposed overlapping site model.**

**The structure also provides details of a possible reaction mechanism. The conformation of bound NADP, with the pro-*S* side of the nicotinamide ring facing toward the protein and the pro-*R* side accessible to solvent, confirms that in the dehydrogenase-catalyzed reaction the hydride is transferred to the pro-*R* side. The observed cluster of conserved residues within the cleft, together with the existing biochemical data, led us to propose the location of a putative active site for the cyclohydrolase reaction. We suggest that a $Y^{52}XXXK^{56}$ motif, identical to the catalytic motif of the short-chain alcohol dehydrogenases and the constellation Ser49–Gln100–Pro102 are important in the cyclohydrolase reaction. These residues are located nearby the predicted position of the folate substrate and are conserved in the bifunctional D/C enzymes, but not in the yeast monofunctional dehydrogenase. The structure provides a good basis for testing the details of the catalytic mechanism of the dehydrogenase/cyclohydrolase family by site-directed mutagenesis.**

## Materials and methods

Crystallization of DC301, the D/C domain (residues 2–301) of the human cytosolic D/C/S enzyme, has previously been described [36]. The three-dimensional structure determination proceeded with the orthorhombic crystal form grown from 21% PEG 3350, 5% glycerol, 0.2M ammonium acetate and 0.1mM sodium citrate pH 5.5. This crystal form belongs to the space group $P2_12_12_1$ with unit cell parameters

**Table 1**

**Data collection and phase determination for DC301.**

| | Dataset | | | | |
|---|---|---|---|---|---|
| | Native 1* | Native 2† | SeMet derivative‡* | SeMet derivative‡* | SeMet derivative‡* |
| **Diffraction data** | | | | | |
| Wavelength (Å) | 1.0716 | 0.9180 | 0.97944 | 0.97917 | 0.96864 |
| Resolution (Å) | 2.0 | 1.5 | 2.5 | 2.5 | 2.5 |
| No. reflections (unique) | 149,470 (35,154) | 753,128 (75,578) | 135,152 (18,138) | 134,029 (18,110) | 131,893 (18,038) |
| Completeness (%) | 90.0 | 83.0 | 89.3 | 89.1 | 88.8 |
| $R_{merge}$ (%) | 3.5 | 6.8 | 4.0 | 5.1 | 4.3 |
| **Phasing** | | | | | |
| No. of sites | – | – | 12 | 12 | 12 |
| Overall <FOM> | – | – | 0.54 | 0.54 | 0.54 |

*Data collected at X4A beamline at BNL.†Data collected at MacCHESS facility. ‡Diffraction data were collected at three different wavelengths for the selenomethionyl derivative of DC301.

a = 67.5, b = 135.8, c = 61.4 Å. There are two molecules in the asymmetric unit and 44% of the volume is taken by the solvent. Crystals were flash frozen by dipping them in heavy mineral oil (N-paratone) and then transferring them directly into the stream of nitrogen at 110K generated by a cryostat (Oxford Cryosystem). Preliminary analysis was performed with diffraction data collected on a RAXIS-IIC area detector mounted on a Rigaku RU-300 rotating anode. The imaging frames were processed and scaled with DENZO and SCALEPACK [37,38] and the data was further analyzed with the CCP4 suite of programs [39].

Screening for heavy-atom derivatives identified one useful platinum derivative with crystals soaked in 9 mM $K_2PtCl_4$ for 19 h. The platinum positions were identified using the technique of Patterson vector superposition included in SHELXS [40]. Four heavy-atom sites were common in 17 of the 20 different vectors used for superposition, with a fifth site present in nine of those. All possible combinations of triplets of the four sites were refined independently using MLPHARE [39] and difference Fourier maps identified the missing fourth site, site number 5 from SHELXS runs and one additional site. These six sites were related by a noncrystallographic twofold axis, consistent with the self-rotation function analysis. An electron-density map calculated with the phases derived from the platinum derivative revealed the location of the protein molecule, but was of insufficient quality to allow tracing of the polypeptide chain.

A selenomethionyl DC301 derivative was expressed, purified and crystallized according to the same protocols as the native protein. Synchrotron diffraction data (see Table 1) were collected at beamline X4A (NSLS, BNL, NY) on a native crystal and a selenomethionyl DC301 derivative in the MAD experimental mode [41]. Selenium positions were found from the anomalous and dispersive Fourier maps calculated with the phases derived from the platinum derivative; no peak was found for the N-terminal selenomethionine residue. The MAD data were analyzed with MLPHARE. These phases were further improved by solvent flattening and histogram matching with DM [39] and gave a readily interpretable map. The model was built with the program O [42] and refined with a combination of simulated annealing (X-PLOR [43]) and maximum-likelihood refinement (REFMAC [39]). SIGMAA-weighted maps [44] were used for model rebuilding. The refinement was initially performed with data extending to 2.0 Å resolution. The five C-terminal residues were not visible in the electron-density map and were not included in the refinement. In molecule A, there was no density for residues 241–250, although the equivalent region has a well-defined density in molecule B. Recently, the diffraction data to 1.5 Å resolution were collected at the MacCHESS facility and the refinement was extended to this resolution. Cycles of refinement and rebuilding continued until convergence was reached. Bulk solvent correction was applied during these calculations. No improvement of the electron density in the above mentioned disordered regions was noticed. No structural outliers were identified by PROCHECK [45]. The final R factor is 0.199 for all data in the range 20–1.5 Å, $R_{free}$ (5% of total reflections) is 0.235 (Table 2).

**Table 2**

**Refinement statistics for DC301.**

| | |
|---|---|
| Resolution | 20.0–1.5 Å (all data) |
| R factor/$R_{free}$* | 0.199/0.235 |
| Rms bond distances | 0.015 Å |
| Rms bond angles | 1.6° |
| $B_{ave}$ | 19.5 |
| Model | Subunit A (residues 2–240, 252–296) |
| | Subunit B (residues 2–296) |
| | 2 molecules of NADP |
| | 407 water molecules |

*R factor = $\Sigma_{hkl} | F_{calc} - F_{obs} | / \Sigma_{hkl} | F_{obs} |$, where the $F_{calc}$ and $F_{obs}$ are the calculated and observed structures factor amplitudes, respectively. $R_{free}$ is calculated with the same formula over a range of selected reflections.

### *Accession numbers*

Atomic coordinates have been deposited in the PDB [46] with the accession code 1A4I. This is NRC publication number 41405.

## Acknowledgements

We thank the people at beamline X4A (Brookhaven, NY) for their precious support. High resolution data were collected at the MacCHESS facility and the support of the staff is gratefully acknowledged. We also thank Narciso Mejia for enzyme preparation and JD Schrag for critical reading of the manuscript. This work was supported by NSERC grant OGP0155357 to MC and MRC grant MT-4479 to REM.

## References

1. Ubbink, J.B. (1994). Vitamin nutrition status and homocysteine: an atherogenic risk factor. *Nutr. Rev.* **52**, 383-387.
2. Butterworth, C.E.J. & Bendich, A. (1996). Folic acid and the prevention of birth defects. *Annu. Rev. Nutr.* **16**, 73-97.
3. MacKenzie, R.E. (1984). Biogenesis and interconversion of substituted tetrahydrofolates. In *Folates and pterins: chemistry and biochemistry of folates.* (Blakley, R. & Benkovic, S., eds), pp. 256-306, John Wiley and Sons, New York.
4. Tan, L.U. & MacKenzie, R.E. (1977). Methylenetetrahydrofolate dehydrogenase, methenyltetrahydrofolate cyclohydrolase and formyltetrahydrofolate synthetase from porcine liver. Isolation of a dehydrogenase/cyclohydrolase fragment from the multifunctional enzyme. *Biochim. Biophys. Acta* **485**, 52-59.

5. Schirch, L. (1978). Formyl-methenyl-methylenetetrahydrofolate synthetase from rabbit liver (combined). Evidence for a single site in the conversion of 5,10-methylenetetrahydrofolate to 10-formyltetrahydrofolate. *Arch. Biochem. Biophys.* **189**, 283-290.
6. Drummond, D., Smith, S. & MacKenzie, R.E. (1983). Methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase from porcine liver: evidence to support a common dehydrogenase-cyclohydrolase site. *Can J Biochem Cell Biol* **61**, 1166-1171.
7. Appling, D.R. & Rabinowitz, J.C. (1985). Evidence for overlapping active sites in a multifunctional enzyme: immunochemical and chemical modification studies on C1-tetrahydrofolate synthase from *Saccharomyces cerevisiae*. *Biochemistry* **24**, 3540-3547.
8. Cohen, L. & MacKenzie, R.E. (1978). Methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/ formyltetrahydrofolate synthetase from porcine liver. Interaction between the dehydrogenase and cyclohydrolase activities of the multifunctional enzyme. *Biochim. Biophys. Acta* **522**, 311-317.
9. Wasserman, G.F., Benkovic, P.A., Young, M. & Benkovic, S.J. (1983). Kinetic relationships between the various activities of the formyl/methenyl/methylenetetrahydrofolate synthetase. *Biochemistry* **22**, 1005-1013.
10. Hum, D.W. & MacKenzie, R.E. (1991). Expression of active domains of a human folate-dependent trifunctional enzyme in *Escherichia coli*. *Protein Eng.* **4**, 493-500.
11. Pelletier, J.N. & MacKenzie, R.E. (1994). Binding of the 2′,5′-ADP subsite stimulates cyclohydrolase activity of human NADP(+)-dependent methylenetetrahydrofolate dehydrogenase/cyclohydrolase. *Biochemistry* **33**, 1900-1906.
12. Pelletier, J.N. & MacKenzie, R.E. (1995). Binding and interconversion of tetrahydrofolates at a single site in the bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase. *Biochemistry* **34**, 12673-12680.
13. Rossmann, M.G., Liljas, A., Brändén, C.-I. & Banaszak, L.K. (1975). Evolutionary and structure relationship among dehydrogenases. In *The Enzymes: Oxidation–Reduction. (*Boyer, P.D., ed), pp. 61-102, Academic Press, New York.
14. Wierenga, R.K., Terpstra, P. & Hol, W.G. (1986). Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **187**, 101-107.
15. Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends. Biochem. Sci.* **20**, 478-480.
16. Birktoft, J.J., Rhodes, G. & Banaszak, L.J. (1989). Refined crystal structure of cytoplasmic malate dehydrogenase at 2.5 Å resolution *Biochemistry* **28**, 6065-6081.
17. Tsigelny, I. & Baker, M.E. (1996). Structures important in NAD(P)(H) specificity for mammalian retinol and 11-*cis*-retinol dehydrogenases *Biochem. Biophys. Res. Commun.* **226**, 118-127.
18. Dessen, A., Quemard, A., Blanchard, J.S., Jacobs, W.R.J. & Sacchettini, J.C. (1995). Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*. *Science* **267**, 1638-1641.
19. Fraser, C.M., *et al*., Kelley, J.M. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-403.
20. Fischer, D. & Eisenberg, D. (1996). Protein fold recognition using sequence-derived predictions. *Protein Sci.* **5**, 947-955.
21. Rost, B., Schneider, R. & Sander, C. (1997). Protein fold recognition by prediction-based threading. *J. Mol. Biol.* **270**, 471-480.
22. Persson, B., Zigler, J.S.J. & Jornvall, H. (1994). A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *Eur. J Biochem* **226**, 15-22.
23. Shannon, K.W. & Rabinowitz, J.C. (1988). Isolation and characterization of the *Saccharomyces cerevisiae* MIS1 gene encoding mitochondrial C1-tetrahydrofolate synthase. *J. Biol. Chem.* **263**, 7717-7725.
24. Bélanger, C. & MacKenzie, R.E. (1989). Isolation and characterization of cDNA clones encoding the murine NAD-dependent methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase. *J. Biol. Chem.* **264**, 4837-4843.
25. D'Ari, L. & Rabinowitz, J.C. (1991). Purification, characterization, cloning, and amino acid sequence of the bifunctional enzyme 5,10-methylenetetrahydrofolate dehydrogenase/5,10-methenyl-tetrahydrofolate cyclohydrolase from *Escherichia coli*. *J. Biol. Chem.* **266**, 23953-23958.
26. Pawelek, P.D. & MacKenzie, R.E. (1996). Methylenetetrahydrofolate dehydrogenase/cyclohydrolase from *Photobacterium phosphoreum* shares properties with a mammalian mitochondrial homologue. *Biochim. Biophys. Acta* **1296**, 47-54.
27. Green, J.M., Matthews, R.G. & MacKenzie, R.E. (1986). Stereochemistry of hydride transfer to $NADP^+$ by methylenetetrahydrofolate dehydrogenase from pig liver. In *Chemistry and Biology of Pteridines. (*Cooper, B.A. & Whitehead, V.M., eds), pp. 901-904, Walter de Gruyter and Co., Berlin.
28. Baker, M.E. (1990). A common ancestor for human placental 17 beta-hydroxysteroid dehydrogenase, *Streptomyces coelicolor* actIII protein, and *Drosophila melanogaster* alcohol dehydrogenase. *FASEB J.* **4**, 222-226.
29. Krozowski, Z. (1994). The short-chain alcohol dehydrogenase superfamily: variations on a common theme. *J. Steroid Biochem. Mol. Biol.* **51**, 125-130.
30. Breton, R., Housset, D., Mazza, C. & Fontecilla-Camps, J.C. (1996). The structure of a complex of human 17β-hydroxysteroid dehydrogenase with estradiol and $NADP^+$ identifies two principal targets for the design of inhibitors. *Structure* **4**, 905-915.
31. Barlowe, C.K. & Appling, D.R. (1990). Isolation and characterization of a novel eukaryotic monofunctional NAD(+)-dependent 5,10-methylenetetrahydrofolate dehydrogenase. *Biochemistry* **29**, 7089-7094.
32. West, M.G., Barlowe, C.K. & Appling, D.R. (1993). Cloning and characterization of the *Saccharomyces cerevisiae* gene encoding NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase. *J. Biol. Chem.* **268**, 153-160.
33. Strong, W.B. & Schirch, V. (1989). *In vitro* conversion of formate to serine: effect of tetrahydropteroylpolyglutamates and serine hydroxymethyltransferase on the rate of 10-formyltetrahydrofolate synthetase. *Biochemistry* **28**, 9430-9439.
34. Pasternack, L.B., Laude, D.A.J. & Appling, D.R. (1992). $^{13}C$ NMR detection of folate-mediated serine and glycine synthesis *in vivo* in *Saccharomyces cerevisiae*. *Biochemistry* **31**, 8713-8719.
35. Pasternack, L.B., Laude, D.A.J. & Appling, D.R. (1994). Whole-cell detection by 13C NMR of metabolic flux through the C1-tetrahydrofolate synthase/serine hydroxymethyltransferase enzyme system and effect of antifolate exposure in *Saccharomyces cerevisiae*. *Biochemistry* **33**, 7166-7173.
36. Allaire, M., Li, Y., Mejia, N.R., Pelletier, J.N., MacKenzie, R.E. & Cygler, M. (1996). Crystallization of the bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase domain of the human trifunctional enzyme. *Proteins* **26**, 479-480.
37. Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*. Sawyey, L., Isaacs, N. & Bailey, S., eds), pp. 56-62, SERC Daresbury Laboratory, England.
38. Minor, W. (1993) *XDISPLAYF Program*. Purdue University,
39. Collaborative Computational Project, No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760-763.
40. Sheldrick, G.M., Dauter, Z., Wilson, K.S., Hope, H. & Sieker, L.C. (1993). The application of direct methods and Patterson interpretation to high-resolution native protein data *Acta Cryst. D* **49**, 18-23.
41. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.
42. Jones, T.A., Zhou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110-119.
43. Brünger, A.T. (1993). *X-PLOR Version 3.1*. Yale University Press, New Haven, CT.
44. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140-149.
45. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.
46. Bernstein, F.C., *et al*., Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
47. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.
48. Bacon, D.J. & Anderson, W.F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics* **6**, 219-220.
49. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.